

BEOWULF SYSTEM SCOPE OF WORK AND SPECIFICATIONS:

The supplier must provide the system described below as a configured, working, turnkey beowulf in a cabinet with MPI/Linux installed. Supplier must demonstrate adequate experience in supporting high performance computing (e.g., floating point applications, compilers, number crunching), and installation/supply of beowulf computers.

While the system may be shipped as several physical components that will require assembly by EPA or its contractors, the offerors shall ensure that all components shipped will adequately result in a completely turnkey system. No additional customization or configuration or minor mechanical components (eg mounting hardware) shall be required to make the system fully operation and ready for use. The cabinet will be already cabled. Cables (e.g., power, network) should have loops/extra length to allow hardware on sliding rails to operate both in normal (retracted) mode and with rails fully extended). Offeror will ship the unit partially disassembled (i.e., cabinet, 2U and 4U units, display, shipped separately), with rack prewired/cabled. Assembly at destination will only require inserting computers into racks and plugging in cables.

Warranty and Technical Support Provisions:

The offeror shall provide a minimum of one year of hardware and system warranty, which shall include technical support services necessary for the initial implementation support, and for problem resolution throughout the warranty period. Technical support shall be delivered between the hours of 9:00 a.m. and 5:00 p.m., EST by the contractor to EPA or its contractors via telephone within three business hours after receiving a request for assistance. Also, during the warranty period, when hardware problems are identified and reported, the requested replacement parts shall be shipped within one business day for overnight delivery to EPA.

Supplier must provide the necessary level of problem resolution support to ensure that hardware and software have been correctly installed and configured.

REQUIRED ITEMS (description below):

- Item 1: two (2) dual athlon master nodes, each with keyboard/mouse.
- Item 2: sixteen (16) dual athlon slave nodes
- Item 3: one (1) 24 port 100Mbps network switch associated cables
- Item 4: one (1) Cabinet and power cables
- Item 5: one (1) display
- Item 6: spare disks, spare fans

All nodes must be able to boot with/without keyboard, mouse or video monitor. All nodes must be able to hot plug the keyboard/mouse and monitor (ie. without a KVM switch). Since hot-plugging keyboards and mice is not in the specification (even though it usually works), the keyboard/mouse will connect through the USB ports of the master/slaves. It is acceptable to use

PS2 keyboards and mice through an external PS2 to USB converter box which is moved by hand. The monitor will be hot-plugged by moving the SVGA connector by hand. All nodes must have a hardware clock that maintains time through powerdown and reboots.

All nodes must have temperature, voltage and fan speed monitoring. Alarms must shutdown and power off the node when these variables are out of safe range. Temperature, voltage and fan speed will be logged to syslog at regular intervals (approximately 1 minute, user set-able) and alarms will be similarly logged.

The software for this monitoring/alarm/poweroff will be supplied in source code form.

All nodes, keyboard/mouse trays, will be rack mounted, on sliding rails. The monitor and keyboards/mice will sit on a table next to the cabinet (table supplied by EPA). The ethernet switch will be on a fixed rackmount.

Networking: All nodes will have an onboard 100Mbps NIC, and be connected via a switch. The two master nodes each require an 2nd NIC for connection to the outside world (in the mobos specified, this 2nd NIC is onboard).

Video cards: Both master nodes will have onboard video with at least 4M memory. Slave nodes will have no video card and will be configured to use a serial port for the console. Only one (1) slave will have a video card. This card will be moved between slave nodes when problems cannot be resolved via the serial console. This video card will only be used in text mode, however for ease of management, it should be the same video chipset/card type as the on-board video in the master node.

Hostname labels: all nodes will be labeled with their hostname on both the front and back of the case. Hostnames are listed below (m1, b1, b2..) - do not include the domainname. Dymo marker tape is sufficient.

Node configuration:

There are two types of nodes, master (2 nodes) and slave (16 nodes).

Standard Hardware:

All nodes shall have:

- * dual SMP Athlon CPUs, minimum MP2000+
- * 1 Gbyte DDR-RAM memory, 266MHz FSB
- * onboard IDE/ATA dual channel drive controller
- * 1 1.44Mbyte floppy drive
- * 1 40Gbyte IDE, 7200rpm disk

Note: This disk will have the system files, will have the ext3 filesystem and will be partitioned as follows:

1st partition - root partition - 512Mbytes
(or 1024 cyls, whichever is smallest). Includes /boot
2nd partition - swap - 2Gbytes
3rd partition - /usr - 10Gbytes
4th partition - /usr2 - rest of disk

All partitions will be mounted with the "noatime" option.

Specific Requirements for Slave nodes:

In addition to the standard node requirements above, each slave node will have:

- * Case: 2U
- * Mobo: Tyan S2466GN (300W power supply is sufficient).
- * Onboard 100Mbps ethernet riser card: 1 x 64 bit riser card for later use with myrinet.

All slaves nodes must boot with the console redirected to a serial port.

In addition, one (1) of the slave nodes will have a video card of the same type/specs as the onboard video on the master nodes. This video card will be used when booting through the serial console fails. This card will be moved to other slave nodes as needed. The node in which this video card is installed will still use the serial port for the console.

Specific Requirements for Master nodes:

In addition to the standard node requirements above, each master node will have:

- * Case: 4U (must take PCI cards without a riser card)
- * Onboard U-160 scsi controller:
- * Onboard video, minimum 4Mbyte video memory
- * Onboard dual 100Mbps ethernet
- * Mobo: Tyan S2468UGN (430W power supply)
- * Extra disk: SCSI 70Gbytes, with single ext3 partition,
mounted as /home. All disks of this size
seem to be 10krpm. Only 7200rpm is needed,
but 10krpm is acceptable.
- * CDRW IDE drive:

KEYBOARD (standard desktop style, not rackmount), with free moving 3-button mouse/mousepad (not trackball or mouse built into keyboard). Keyboard/mice will sit on table beside cabinet. Keyboard/mouse must hot-plug into master/slaves. Use of PS2 keyboard/mouse with external PS2/USB converter is acceptable. Includes set of extension cables to plug keyboard/mouse into any node in rack. Keyboard/mouse extension cables must be able to reach from node at top of cabinet, from the bottom of the cabinet for at least 6 feet.

Serial cable/nullmodem/LED mini-tester adapter for connection to slave nodes to monitor serial console: Cable must be of sufficient length to connect to from master nodes to any slave nodes in rack. The LED mini-test adapter is continuously inline and has 7 bipolar red/green LEDs for monitoring activity on the serial line (looks like a null modem). One (1) set for each master node (total of 2).

Network Switch:

One (1) 24 port 100Mbps switch: unmanaged, rackmounted Netgear FS524NA. Rackmount can be 2 L-brackets. No vertical faceplate is needed. Switch will be mounted on the back of the cabinet, with ethernet connectors facing backwards.

Network/Power Cabling:

Supplier will provide network (ethernet) cabling. Cabling will be pre-installed in cabinet by offerer. Generous cable length should be used to allow for possible post-delivery changes in configuration.

While the cables should be secured where they are passing down the sides of the cabinet, enough free cable should be left at the ends where they attach to the nodes/switch, so that the cables can be disconnected and reconnected easily.

All LAN cables will be of the same length. Cables will be of sufficient length to connect any two devices in rack. Loops at node end of cable will be sufficient to allow hosts to operate both in normal position and with the nodes fully forward extended on rails. A loop should be left at the switch end (about 1 foot excess hanging from the RJ-45 connectors).

LAN cables:

24 cables, color red for 100Mbps ethernet communication between slaves, master and switch. All LAN cables are the same length.

WAN cables:

4 cables, color yellow for connecting 2nd ethernet interface on master nodes to outside world. Length 25 feet.

(Not all cables will be used in the initial setup, some are spares).

Configuration information:

Node names, IPs and routing:

There are two networks -

WAN - each master node has a 2nd 100Mbps NIC to the outside world;

LAN - all nodes have a 100Mbps NIC connecting to a switch on the LAN.

WAN: The two master machines have a NIC facing the outside world. The domainname will be nesc.epa.gov. The hostnames will be:

NODE	hostname	IP
master	castor	10.0.1.1
master2	pollux	10.0.1.2

LAN: The two masters and the slave nodes all have a NIC connecting to the switch on a LAN. There is no need for a LAN domainname. The hostnames, IPs will be as follows:

NODE	hostname	IP
master	m1	192.168.1.1
master2	m2	192.168.1.2
node1	b1	192.168.1.129
node2	b2	192.168.1.130
.		
.		
node16	b16	192.168.1.144

The /etc/hosts file will be

```
#-----/etc/hosts-----
10.0.1.1          castor.nesc.epa.gov      castor
10.0.1.2          pollux.nesc.epa.gov     pollux
192.168.1.1       m1
192.168.1.2       m2
192.168.1.129     b1
192.168.1.130     b2
.
.
192.168.1.144     b16
#-----
```

The slaves only need a route to the LAN (i.e., no default route to the WAN).

Display:

15" CRT monitor 1024x768@60Hz, beige case. Display will sit on a desk beside the cabinet. Includes extension cord to plug monitor into any node in rack, incase slaves cannot be booted through serial port.

display extension cord must be able to reach from node at top of cabinet, and extend 6 feet beyond bottom of cabinet. Power cord extension will be long enough to connect to power strip in cabinet (power cord should be able to reach 6 feet beyond cabinet bottom).

Software Configuration, System setup:

All nodes installed with RedHat Linux 7.2,

MPICH (latest version) will be installed on the two (2) master nodes at /usr/local/mpich_version_number.

This directory will be available as a link from /usr/local/mpich on the master nodes.

On both master machines, /usr/local/mpich will be nfs exported (rw,link_absolute,insecure,no_root_squash) to all machines on the LAN (ie to 192.168.1.0/24).

The slave machines will nfs mount (ro) from the primary master m1 m1:/usr/local/mpich at /usr/local/mpich.

The exports from the backup master machine m2 will not be mounted on any machine (directories from the backup master m2, will be mounted manually at customer site later only on failure of machine m1).

MPICH will use rsh (rather than ssh) for process initiation and be setup to use gcc/g77/g++ compilers. (We _do_not_ want Scyld.)

The /home directory on both master machines will be nfs exported to the LAN (ie to 192.168.1.0/24) as (rw,link_absolute,insecure,no_root_squash)

The slaves will mount the disk from master m1 only as /home (rw). (later at the customer site, on failure of machine m1, the sysadmin will manually mount /home from the backup master m2 on the slaves).

The two master machines will have ftpd running. All machines on the LAN will be able to transfer files to/from the master nodes by anonymous ftp. There will be no ftpd running on the slave nodes.

The two master machines will use 134.67.208.5, 134.67.208.10 for DNS.

Telnet login will be enabled between all machines.

ntp will run on all nodes. All nodes will be ntp peers. an ntp.conf file that will work satisfactorily on all nodes could be

```
#-----/etc/ntp.conf-----  
driftfile /etc/ntp.drift  
server 127.127.1.0  
fudge 127.127.1.0 stratum 13  
peer m1  
peer m2  
peer n1  
.  
.  
peer n16  
#-----
```

Cabinet: full height (42U)

Must be accessible from the back, for attending to cabling. Will accept computers in sliding trays, includes mounting hardware (nuts/screws). With doors, 3x500 CFM fans and on casters.

Power strips must be able to handle a fully populated cabinet (approx 20 units requiring power). Power requirements for cabinet must be handled by a single cord with an L5-30P plug exiting the bottom of the cabinet. Power cord must extend at least 5 feet beyond bottom of cabinet.

Cabinet must meet industry standards for safety and mounting hole configuration, and include the following:

Locks on doors

Adjustable mounting channels

Doors: front door perforated, (solid rear cabinet wall)

Cabinet finish: textured, semi-gloss powder coat

Hang points or clips will be provided in cabinet to route all cables. All cables must be of sufficient length to be routed through these hang points (kdb, mouse, serial, display, power, network).

Position of nodes:

The cabinet is almost full (all but 2U used by nodes). Nodes will be mounted from the top in the order n1..n8,m1,switch (mounted in back, taking up 2U of space),m2,n8..n16.

Spares:

Four (4) 40GB ATA/100 7,200RPM IDE Hard Drive drives will be formatted and have the same content as one of the slave nodes.

Four (4) CPU fans four (4) chassis fans for the slave nodes

Two (2) chassis fans for the master nodes